



Influence of talker discontinuity on cortical dynamics of auditory spatial attention

Mehraei, Golbarg; Shinn-Cunningham, Barbara; Dau, Torsten

Published in:
NeuroImage

Link to article, DOI:
[10.1016/j.neuroimage.2018.06.067](https://doi.org/10.1016/j.neuroimage.2018.06.067)

Publication date:
2018

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Mehraei, G., Shinn-Cunningham, B., & Dau, T. (2018). Influence of talker discontinuity on cortical dynamics of auditory spatial attention. *NeuroImage*, 179, 548-556. <https://doi.org/10.1016/j.neuroimage.2018.06.067>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Accepted Manuscript

Influence of talker discontinuity on cortical dynamics of auditory spatial attention

Golbarg Mehraei, Barbara Shinn-Cunningham, Torsten Dau

PII: S1053-8119(18)30574-3

DOI: [10.1016/j.neuroimage.2018.06.067](https://doi.org/10.1016/j.neuroimage.2018.06.067)

Reference: YNIMG 15073

To appear in: *NeuroImage*

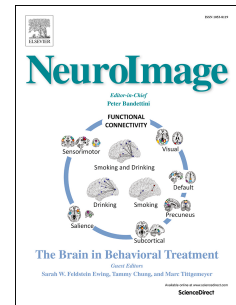
Received Date: 5 January 2018

Revised Date: 12 June 2018

Accepted Date: 25 June 2018

Please cite this article as: Mehraei, G., Shinn-Cunningham, B., Dau, T., Influence of talker discontinuity on cortical dynamics of auditory spatial attention, *NeuroImage* (2018), doi: 10.1016/j.neuroimage.2018.06.067.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Influence of talker discontinuity on cortical dynamics of auditory spatial attention

*Golbarg Mehraei^{a,1}, Barbara Shinn-Cunningham^{b,c,1}, Torsten Dau^{a,1}

**Corresponding author: Golbarg Mehraei*

*Ørsted's Plads Building 352 Kongens Lyngby, 2800, Denmark
email: gmehraei@alum.mit.edu*

*^aHearing Systems Group, Technical University of Denmark, Ørsted's Plads Building 352,
2800, Kongens Lyngby, Denmark*

*^bCenter for Research in Sensory Communication and Emerging Neural Technology, Boston
University, Boston, Massachusetts, 02215*

^cDepartment of Biomedical Engineering, Boston University, Boston, Massachusetts, 02215

Abstract

In everyday acoustic scenes, listeners face the challenge of selectively attending to a sound source and maintaining attention on that source long enough to extract meaning. This task is made more daunting by frequent perceptual discontinuities in the acoustic scene: talkers move in space and conversations switch from one speaker to another in a background of many other sources. The inherent dynamics of such switches directly impact our ability to sustain attention. Here we asked how discontinuity in talker voice affects the ability to focus auditory attention to sounds from a particular location as well as neural correlates of underlying processes. During electroencephalography recordings, listeners attended to a stream of spoken syllables from one direction while ignoring distracting syllables from a different talker from the opposite hemifield. On some trials, the talker switched locations in the middle of the streams, creating a discontinuity. This switch disrupted attentional modulation of cortical responses; specifically, event-related potentials evoked by syllables in the to-be-attended direction were suppressed and power in alpha oscillations (8-12 Hz) were reduced following the discontinuity. Importantly, at an individual level, the ability to maintain attention to a target stream and report its content, despite the discontinuity, correlates with the magnitude of the disruption of these

cortical responses. These results have implications for understanding cortical mechanisms supporting attention. The changes in the cortical responses may serve as a predictor of how well individuals can communicate in complex acoustic scenes and may help in the development of assistive devices and interventions to aid clinical populations.

Keywords: auditory attention | event-related potentials | neural oscillations | alpha lateralization

Abbreviations: electroencephalography (EEG); Event-related Potential (ERP); Consonant Vowel (CV); Interaural timing difference (ITD); Finite Impulse Filter (FIR); Attentional Modulation Index (AMI); Mismatch Negativity (MMN); Event-related Desynchronization (ERD)

1. Introduction

Attention plays a fundamental role in understanding complex auditory scenes, operating as a form of sensory gain-control that directly alters the representation of information in the cortex. Specifically, magnetoencephalography and electroencephalography (EEG) studies have shown that selective auditory attention directly modulates event-related potentials (ERPs) evoked by sounds and generated by neural activity in auditory cortex (Hillyard et al., 1973, Picton and Hillyard, 1974, Chait et al., 2010, Ding and Simon, 2012, Choi et al., 2014): ERPs of attended sounds are enhanced while the ERPs of distractor sounds are suppressed (Choi et al., 2014). The degree of modulation of ERPs correlates with individual differences in performance in auditory selective attention tasks (Choi et al., 2014, Dai and Shinn-Cunningham, 2016), suggesting a strong link to perception.

Selective auditory attention also influences ongoing neural alpha oscillations (8-12 Hz) (Strauß et al., 2014, Wöstmann et al., 2015, 2016), which are linked to inhibition of the processing of task-irrelevant information (Thut et al., 2006, Klimesch et al., 2007, Wöstmann et al., 2015). Attentive focusing to one side in auditory space leads to a relative decrease in alpha power in contralateral compared to ipsilateral brain regions (Frey et al., 2014) and governs success of selective attention, isolating one stimulus at a specific spatial location (Kerlin et al., 2010).

Although much effort has been put into studying the relationship between the neural processes controlling attention and auditory scene analysis, little work has gone into understanding how perceptual discontinuities in acoustic scenes affect the neural processing of sustaining auditory attention. In a classical "cocktail party", talkers can change location or a conversation may jump from one speaker to another. These perceptual discontinuities of acoustic features, such as in talker or location, have been shown to affect our behavioral ability

to maintain attention to sound streams, even when the discontinuous feature is not the focus of attention (Best et al., 2008, 2010, Maddox and Shinn-Cunningham, 2012, Bressler et al., 2014).

Here, we investigated how perceptual discontinuity of the talker affect the cortical processes responsible for focusing auditory spatial attention. We analyzed changes in ERP magnitudes and alpha power. EEG recordings showed that when listeners are attending to a particular location, a switch in talker disrupts ERP modulation and decreases power in the alpha band. In addition, the lateralization of alpha power with respect to the side of attention is disrupted following the perceptual discontinuity in talker. Critically, at an individual level, the magnitude of the suppression in ERPs and alpha power predicts how well a listener maintains attention and recalls the attended stimuli, showing a direct link between these neural markers and perceptual outcome.

2. Materials and Methods

2.1. Apparatus

All measures were obtained with subjects seated in an acoustically and electrically shielded booth (double-walled IAC booth, Lyngby, Denmark). A desktop computer outside the booth controlled all aspects of the experiment, including triggering, sound delivery and storing data. The stimuli were presented via Fireface UCX (RME, Haimhausen Germany) and triggers were sent from a RME ADI-8 trigger box (RME, Haimhausen Germany). A headphone driver presented sound through ER-2 insert phones (Etymotic, Elk Grove Village, IL). All sounds were digitized at a sampling rate of 44.1 kHz. During the active portion of the EEG experiment, the subjects responded using the numerical pad on a keyboard.

2.2. Subjects

Nineteen young (median = 25 y; range = 22-34 y; 5 females) right-handed listeners took part in this study. All subjects had pure-tone thresholds below 20 dB hearing level (HL) at octave frequencies between 0.25 and 8 kHz. The subjects provided written informed consent and were financially compensated for their participation. Informed consent was obtained in accordance with protocols established at Technical University of Denmark.

2.3. Stimuli

Stimuli consisted of consonant vowel syllables (CVs) of \ba\, \da\, or \ga\ spoken by a native English male and female talker. CVs were recorded in a sound-proof booth with a large diaphragm condenser microphone (AudioTechnica AT4033, Stow, OH, USA) through a Duet analog-to-digital interface (Apogee Electronics Corp., Santa Monica, CA, USA) at a sampling

rate of 44.1 kHz at 16-bit resolution. Sound files were edited on the digital audio workstation, Digital Performer 7 (MOTU, Cambridge, MA, USA). Auditory materials were presented at an average intensity of ~ 70 dB sound pressure level (SPL).

For each trial, an initial 0.1 s broadband noise was presented diotically to serve as a normalization factor for inherent individual differences in overall ERP magnitude. The noise was ramped with a $0.02 \text{ s } \cos^2$ rise-decay to minimize the use of onset cues. Following the noise-burst, two spatially separated isochronous streams of CV syllables were presented: one from the left (ITD of -0.028 s , corresponding to roughly -30° azimuth), and one from the right (ITD of 0.028 s , $+30^\circ$). Five CV syllables were randomly chosen for each auditory stream with the constraint that the same CV could not be presented simultaneously across the two auditory streams. Each CV was zero-padded at the end such that the overall duration was 0.388 s . Additionally, each CV syllable was ramped with a $0.02 \text{ s } \cos^2$ rise-decay to minimize spectral splatter. As shown in Fig. 1C, by design, the timing of the CVs in the two locations was offset in time to allow isolation of the ERPs evoked by each CV. The leading stream, always the target in the experiment, started 0.6 s after the onset of the noise-burst. The lagging auditory stream started 0.18 s after the onset of the leading stream. The inter-stimulus interval (offset to onset) within each stream was 0.045 s . The initial talkers in the left and right auditory streams were randomly selected with equal probability from trial to trial.

2.4. Procedure

The experiment consisted of both passive and active listening conditions. Passive and active conditions were performed in separate blocks. In the passive listening condition, participants watched a silent, captioned movie of their choice, ignoring the acoustic stimuli.

In the active portion of the experiment, participants fixated on a centrally presented dot. As shown in Fig. 1A, at the start of each trial, a visual cue of a left or right arrow was presented, indicating the to-be-attended side; 0.5 s after the cue onset, there was a 1 s fixation period after which the stimulus was presented. Approximately 0.2 s after the offset of the last CV in the stimulus, a circle appeared around the fixation point, indicating the response period. After a 2 s long response time, the circle changed colors to provide feedback: green to indicate a correct response or red to indicate an incorrect response, respectively. Approximately 1 s (jittered $0.99\text{--}1.01 \text{ s}$) after the response period, the next trial began.

Subjects were instructed to count and report the number of /ga/ syllables they heard in the cued target stream, ignoring the switch in talker if it occurred in the trial. The number of /ga/ syllables on any trial could vary between 0-5. On average two /ga/ syllables were presented. More trials contained a lower number of /ga/ syllables (0-2); the percentage of the trials for 0-5 /ga/ syllables was approximately 14.7%, 34.5%, 31.6%, 15.7%, 3%, and 0.3%, respectively.

On half the trials, a discontinuity was introduced in the task-irrelevant acoustic feature: the talkers swapped locations in the third CV presentation. This is referred to as a “switch

trial". On the other half of the trials, the talker in each location remained the same, referred to as a "continuous trial". Statistically identical stimuli were presented to participants during the passive listening condition. Each participant performed 132 trials for each condition. The trial order was fully randomized.

Including preparation time, the experiment lasted approximately 2h. Prior to the experiment session, each subject had approximately an hour long training session. The training was completed when listeners reached a performance score of 70% trial correct on the continuous trials, well above the chance level of 17%. All but one of the participants were able to reach this criterion; the remaining subject, who reached a performance level of 68%, did not perform the main experiment.

2.5. EEG Data Recording and Analyses

Cortical responses were recorded using a 32-channel EEG system (Biosemi Active 165 II system, Amsterdam, Netherlands) at a sampling rate of 2048 Hz. Two additional electrodes were placed on the mastoids for reference and another four electrodes were placed around the eyes to monitor eye movement.

For EEG data analyses, we used the Fieldtrip toolbox (Oostenveld et al., 2011), EEGLab toolbox (Delorme and Makeig, 2004) and customized Matlab scripts. Continuous data were re-referenced to the average mastoids, highpass-filtered at 1 Hz (1408th order windowed sinc finite impulse response filter, FIR; zero-phase lag), and lowpass-filtered at 20 Hz (1408th order windowed sinc FIR; zero-phase lag). Independent component analysis was used to reject components corresponding to eye blinks and saccadic eye movements. For the ERP analysis, data were down-sampled to 256 Hz and epoched from -0.2 to 3.2 s relative to the onset of the initial noise burst in the trial. Epochs were rejected if the mean amplitude of a trial was a standard deviation or more away from the mean of the distribution across trials. Trials were grouped into two types, continuous and switch trials. To fairly compare across listeners, we used the first 98 remaining trials after the rejection from each condition.

Spectral analysis ($t=0-3.2$ s) was performed using the original sampling rate (2048 Hz). For each electrode, the induced (i.e., average evoked response subtracted from each trial) spectral power and time-frequency content were estimated using the multi-taper method (Thomson, 1982). By removing the averaged evoked response in the spectral analysis, we could analyze the effect of a switch on the spectral power independently from any effect observed in the ERP. Three bi-orthogonal prolate-spheroidal sequences were used in this method to minimize the spectral leakage outside of the bandwidth of 1.33 Hz (Slepian, 1978). A moving window of 0.28 s with a step-size of 0.05 s was used for the computation of the time-frequency representation of induced alpha power. Because alpha frequency varies from subject to subject (Nunez et al., 1978), we determined the individual alpha frequency on a subject basis, defined as the frequency between 8-12 Hz with maximum power (Klimesch, 1999). Using this

subject-specific frequency, we defined each individual alpha band as 2 Hz above and below this peak. To compute the across-subject average induced alpha power, we averaged across these subject-specific alpha bands.

2.6. Attention Indices

Two indices of attentional modulation of neural responses were calculated: amplitude analysis of the N1 of the ERP and the attentional modulation index of induced alpha power (AMI α ; (Wöstmann et al., 2016)). For the ERP analysis, the amplitude of the N1 component was calculated from the individual-subject average ERPs for each electrode, computed by finding the local minimum within a fixed time window positioned from 0.1-0.2 s after each CV onset. For each listener, the N1 in the six front-central electrodes (F3, F4, FC1, FC2, Fz and Cz), which yielded the strongest auditory-evoked responses (Fig. 1D), were averaged together. Inherent individual differences in overall ERP magnitude were large on an absolute scale. We therefore normalized (division) each individual subject's ERPs with the amplitude of the N1 response to the noise-burst at the start of each trial, **averaged over all conditions**. We quantified how the N1 is modulated by attention by comparing the N1 peak amplitudes of each CV in the target stream across conditions (i.e., passive vs. active condition, continuous vs. switch trial).

The AMI α , $[AMI\alpha = (\alpha_{left} - \alpha_{right}) / (\alpha_{left} + \alpha_{right})]$, revealed a spatially resolved measure of attentional effects on alpha power (8-12 Hz) at each electrode. For each condition, trials were separated into attend left and right. The alpha power for each channel (32 channels) in attend left and attend right trials were analyzed separately in two time windows to determine the alpha power before ($t=0.6-1.466$ s) and after a discontinuity ($t=1.467-3.2$ s). The AMI α was computed for each of these two windows.

2.7. Statistical Testing

Unless otherwise specified, statistical inference was performed by fitting linear regression models to the data and adopting a model comparison approach (Baayen et al., 2008). Fixed-effects terms were included for the various experimental factors whereas subject-related effects were treated as random. In order to not over-parameterize the random effects, models were compared with and without each term using the Akaike information criterion (Pinheiro and Bates, 2000). All model coefficients and covariance parameters were estimated using restricted maximum likelihood as implemented in the lme4 library in R. An F approximation for the type-II scaled Wald statistic was employed to make inferences about the fixed effects (Kenward and Roger, 1997): this approximation is more conservative in estimating Type I error than the Chi-squared approximation and performs well even with complex random-effects covariance structures (Schaalje et al., 2002). The p-values and F-statistics based on this approximation are reported.

When testing for differences in mean results, we applied parametric t-tests when the data conformed to normality assumptions ($p > 0.05$ in Shapiro-Wilk test) and non-parametric Wilcoxon signed rank test otherwise. Z and P-values are reported for Wilcoxon rank test. For correlation analyses we used the Spearman correlation. Multiple comparisons were corrected using the false discovery rate to limit Type I error.

3. Results

3.1. Switching of talker reduces behavioral performance

Fig. 2A compares the percent correct responses in trials where the talker in the target location remained the same (i.e., the continuous trials) and where it switched (i.e., the switch trials). When the task-relevant feature (location) and the task-irrelevant feature (talker) were both continuous in the target stream, average performance across subjects was 86.6% correct. However, when the talkers at the target and distractor locations switched, performance dropped significantly, to 71.4% correct (Wilcoxon signed-rank test; $z = 3.82$; $p < 0.001$).

To determine whether target position influenced error rate, we computed the percentage of errors made as a function of target CV position in trials with only a single target (Fig. 2B). We limited our analysis to trials with only single target CVs because the error rates in trials with multiple targets are not independent from one position to another. There was a non-significant trend of which the largest errors in the switch trials occurred when the trial contained the target /ga/ CV at the time of the switch (Fig. 2B, red). Linear mixed-effect regression model of the error rates, with both trial type and CV position and their interaction as regressors, showed significant main effect of position of the target CV ($F_{(4,162)} = 3.65$, $p = 0.007$). There was no significant main effect of trial type or its interaction with target position. The lack of effect of trial type on behavioral performance does not suggest that the switch has no significant effect on the performance because only 34.5% of overall trials were included in this analysis. It is likely that trials with >1 target CV are more demanding and the switch has more of a detrimental effect. Indeed, within the switch trials, about 35% of the errors occurred in the trials with 2 target CVs compared to the single target CV trials that had an error rate of 25%. Nevertheless, when pooled across all trials, the effect of the switch is apparent as shown in Fig. 2A.

3.2. Attention modulates ERPs

The normalized ERP N1 amplitudes, typically occurring ~ 0.1 - 0.15 s after syllable onsets, were calculated separately for each subject, CV, and attentional condition (Fig. 3C). For the same physical stimuli, N1 magnitudes differ between active (Fig. 3C, filled boxes) and passive listening conditions (Fig. 3C, open boxes). Specifically, compared to the evoked responses in the passive listening condition, in the active listening conditions, N1s for CVs

in the to-be attended target stream are enhanced (i.e., increased negativity; see Table 1 for statistical summary). A linear-mixed effect regression model of the ERP amplitudes with CV position and attentional condition (passive vs. active) as regressors yields a significant effect of attentional condition ($F_{(1,313.15)} = 26.69, p < 0.001$) and CV position ($F_{(4,307.94)} = 42.9, p < 0.001$). There was no significant interaction. We also observed a suppression of the N1s for the CVs in the distractor stream. However, a statistical analysis was not performed on the distractor stream because the N1s were difficult to identify in the active listening condition, even though they were clearly identifiable in the passive condition.

3.3. Talker discontinuity disrupts attentional modulation of ERPs

As expected, comparison of the N1s for the continuous (Fig. 3A, black trace) and switch active trials (Fig. 3A, red trace) showed no significant difference in N1 amplitude before the switch in talker. At the time of the switch (yellow highlighted region in Fig. 3A), there was an enhancement of the N1 response relative to when there was no switch in talker. Immediately following this discontinuity, there was an observed suppression of the N1 to the subsequent target CV, as seen in the blue highlighted region in Fig. 3A and C ($z=2.73, p=0.003$). This observation is confirmed with a linear-mixed effect regression model of the ERP amplitudes with CV position and trial type (continuous vs. switch) as regressors. The model yields a significant effect of position ($F_{(4,131.79)} = 22.56, p < 0.0001$) and interaction of position and trial type ($F_{(4,131.16)} = 3.22, p = 0.015$). There was no significant main effect of trial type. The suppression of the N1 following the switch was transient; the N1 to the last CV (i.e., ~1 s after the switch) did not show this suppression.

To confirm that the observed reduction in the N1 following the discontinuity is linked to attention, we compared continuous and switch trials in the passive condition (Fig. 3B). The corrected multiple comparisons showed a significant enhancement of the N1 at the time of the switch ($z=2.82, p=0.02$), the mismatch negativity (MMN), indicating the deviance in the stream. However, we found no notable difference in the N1 of the leading stream following the switch (Fig. 3B). This suggests that the reduction observed following the switch in the active listening condition was likely related to attention as it was not observed in the passive condition.

3.4. Change in alpha power with talker discontinuity

We computed how talker discontinuity affected induced alpha neural oscillations, which are thought to play a functional role in inhibiting processing of task-irrelevant information (Klimesch et al., 2007, Wöstmann et al., 2016). As seen in Fig. 4, an across-condition comparison of all 32 channels showed a significant reduction of induced alpha power following a switch in talker (t-test with false discovery rate correction, $t=3.39, p<0.05, df=18$). Decreased power in the alpha band occurred between the time window of 1.79-2.37 s, coinciding

with the reduced N1 amplitude. The decrease in power was largest in the parietal and occipital channels, as shown in the scalp topography in Fig. 4, consistent with a parietal generator.

The effect of talker discontinuity on the neural representation of attended location was quantified by calculating the attentional modulation index of induced alpha power (AMI α) for all 32 channels during stimulus presentation. Trials for each condition were separated into attend left and attend right trials. AMI α was computed as a response $(\alpha_{left} - \alpha_{right}) / (\alpha_{left} + \alpha_{right})$ for time windows before and after the switch. A positive AMI α indicates larger neural responses for attention-left trials and negative AMI α indicates larger responses for attention-right trials. A difference of the AMI α between the left and the right hemispheres indicates a hemispheric lateralization of neural responses due to focus of spatial attention.

As shown in Fig. 5, in the time window before the switch, the mean AMI α was positive at channels over the left hemisphere but not significantly different from zero over the right hemisphere. This asymmetry is likely related to the asymmetric representation of spatial information in brain regions, including parietal cortex. Specifically, regions in the left cortex primarily represent contralateral (right) exocentric space, while regions in the right hemisphere dominantly represent left (contralateral) exocentric space, but also right exocentric space (Kaiser et al., 2000, Huang et al., 2014).

Within the continuous and switch trials, AMI α was significantly different between left and right hemispheres before a potential switch in talker (Fig. 5A and B; one-tailed paired t-test, $t=2.97$, $p=0.004$; $t=3.47$, $p=0.001$, $df=18$). As expected, there was no significant difference in the lateralization of alpha across trial types (i.e., continuous vs. switch trials) in this time window ($t=-0.03$, $p=0.513$, $df=18$). However, we found that the lateralization of the AMI α was significantly higher in the continuous than in the switch trials in the time window following a potential switch ($t=2.27$, $p=0.018$, $df=18$): in the continuous trials, where the talker in the attended location stayed the same, AMI α remained significantly lateralized (Fig. 5A; $t=1.88$, $p=0.039$, $df=18$) but the lateralization of the AMI α was disrupted when the talker switched location (see the topography in Fig. 5B; $t=0.37$, $p=0.358$, $df=18$).

3.5. Changes in neural response correlate with behavioral performance

We observed individual differences not only in behavioral performance but also in the magnitude of N1 modulation and alpha power changes with a discontinuity in talker. We tested whether the differences observed in the neural responses predicted a listener's ability to maintain attention on a sound stream when the talker is discontinuous. We compared the magnitude of the decrease in both N1 and induced alpha power following a discontinuity in talker to the degree to which this discontinuity affected behavioral performance (i.e., the difference in performance between switch and continuous trials). We found significant correlations between the behavioral cost and both the suppression of the N1 (Fig. 6A; $r=-0.61$, $p=0.005$) and the decrease in alpha power (Fig. 6B; $r=0.53$, $p=0.02$) following the switch in talker.

Specifically, listeners whose performance was degraded more by talker discontinuity showed a larger decrease in both neural responses following the switch.

4. Discussion

Here we showed that discontinuities that may be encountered in everyday acoustic scenes disrupt cortical processing involved in selecting and maintaining attention, thereby affecting perception. Specifically, a change in talker from an attended location reduced behavioral performance. Following this change, there was a reduction in N1 amplitude evoked by a subsequent target syllable and a decrease in alpha power, associated with suppression of distractor syllables. The magnitude of the decreases in both N1 amplitude and induced alpha power predicted the behavioral cost associated with the perceptual discontinuity. Ordinarily, focused spatial attention is associated with strong lateralization of alpha power (enhanced alpha contralateral to the distractor stimuli) (Frey et al., 2014, Wöstmann et al., 2015). Interestingly, following the switch in talker, the hemispheric lateralization of alpha was disrupted, yielding a diffuse pattern across the scalp. To our knowledge, this is the first study that has demonstrated this neural correlate of disruption of auditory attention.

Past behavioral studies have shown that discontinuity of an unattended/task-irrelevant feature impairs one's ability to selectively attend to a sound stream (Maddox and Shinn-Cunningham, 2012, Bressler et al., 2014). In these studies, when the unattended feature was discontinuous (e.g., switching talkers in the attended location), listeners were more likely to report content from a competing syllable that matched the preceding target in its irrelevant feature (i.e., report information from the same talker but from the wrong location rather than the information from the new talker in the to-be-attended target location; Maddox and Shinn-Cunningham (2012)). These results show that even when a feature should be ignored to perform the task as instructed, its continuity has an obligatory influence on selective auditory attention. Consistent with this previous work, we found a significant decrease in performance when listeners were supposed to attend to location regardless of talker identity, but the talker at the attended location switched identities. It may be more natural to attend to a talker rather than a location; however, the same behavioral effects have been observed when attending to a talker that moves in space (Maddox and Shinn-Cunningham, 2012).

While there is an effect of perceptual discontinuity on behavioral performance, until now, it was not clear how this affects the cortical control of attention. When listeners need to analyze the spectrotemporal content of a sound source in the presence of simultaneous, competing sources, they must sustain selective attention on the target source. In such situations, attention has a substantial effect on the sensory representation of the sound mixture in the cortex. Consistent with past work, we found that attention enhanced N1s evoked by CVs in the target stream (Picton and Hillyard, 1974, Choi et al., 2013, 2014). We also observed that the N1s evoked by CVs in the distractor/unattended stream were suppressed (relative

to the passive condition), suggesting that auditory attention operates as a form of sensory gain-control (see also Choi et al. (2014)).

When the talkers at the attended and ignored locations switched, the effects on the neural response were two-fold: there was 1) an enhancement of the N1 evoked by the first CV following the switch and 2) a suppression of the N1 evoked by the subsequent CV following the change (Fig. 3A). The enhancement of the N1 evoked by the third CV in the target stream is consistent with the MMN response associated with a deviance in the stream (i.e., a change in talker). Consistent with the fact that mismatch negativities are pre-attentive, the MMN was also observed in the passive condition (Fig. 3B). Thus, the enlarged response to the third CV response is likely not linked to attention, but rather represents an automatic response to deviations from expectations in sound streams (Näätänen et al., 1978). In contrast, following this enhancement, the N1 evoked by the fourth target CV had a significantly reduced amplitude (Fig. 3A). This was not observed in the passive trials (Fig. 3B), suggesting that this effect reflects a disruption of cortical mechanisms of attention that lead to target enhancement. Although we cannot infer much about the N1 at the time of the switch, as it overlaps with the MMN, the suppression of the N1 following the switch seems to reflect a degradation of the sensory representation of that target CV in the cortex, which interfered with extracting target content. The attentional modulation of N1 recovered about 1 s after the discontinuity, as seen in the N1 amplitude evoked by the last CV in the target stream. Future work may utilize this ERP method to investigate whether the recovery of attention is prolonged in older and/or hearing-impaired listeners following perceptual discontinuities, as some evidence suggests longer neural recovery times and slowing of cognitive processing associated with age (Schneider and Pichora-Fuller, 2001, Lu et al., 2011).

Along with the suppression of the N1 following the talker discontinuity, the power in the alpha band (8-12 Hz) decreased (Fig. 4). This event-related desynchronization (ERD) persisted through several cycles of the alpha oscillations and occurred around the time at which the third CV in the target stream was presented. It is possible that the alpha desynchronization and N1 effects are linked: previous work has found that phase-locked alpha and theta oscillations generate the ERP N1-P2 complex (Klimesch et al., 2004). However, we analyzed induced alpha power (averaged evoked response removed). Although one might expect that the magnitude of alpha power, which is associated with suppression of distractors, is related to the degree to which the N1 amplitude is modulated by attentional state, we found no significant relationship between these neural measures. Although this negative result cannot be interpreted as support for the null hypothesis (that alpha modulation and N1 modulation are independent), this negative result calls for further investigation into whether or not there is a direct relationship between alpha strength and N1 suppression. Our interpretation of the ERD in the alpha band is based on its functional role in the inhibition of task-irrelevant information (Thut et al., 2006, Klimesch et al., 2007, Wöstmann et al., 2015): following the

discontinuity in talker, the suppression of power in the alpha band suggests that the cortical mechanisms responsible for inhibiting the distractor stream were disrupted.

Alternatively, this desynchronization of alpha may reflect the increase in attentional demand following the discontinuity (Dujardin et al., 1993). However, if the change in the alpha power was indeed reflecting task engagement, we would not expect to see differences in the lateralization of induced alpha across continuous and switch trials (Fig. 4), where the effect of task engagement is removed through the difference metric used here. Moreover, although this condition was not included here, we did not observe a decrease in induced alpha power following a discontinuity when listeners are instructed to attend to the talker, regardless of the location (See supplementary material). If the effect we observe in Fig. 4 was due to task engagement, it should be present in both attend-talker (not reported here) and attend-location conditions.

In this spatial attention task, alpha power lateralization depended on the direction to which attention was directed (Fig. 5; Kerlin et al. (2010), Wöstmann et al. (2016)): alpha power tended to increase in the hemisphere ipsilateral to the exogenous locus of attention and decrease in the hemisphere ipsilateral to the side that subjects ignored. This pattern was most obvious in the posterior channels, consistent with activity in parietal regions (Colby and Goldberg, 1999, Smith et al., 2010, Michalka et al., 2015). The pattern unlikely reflects the effects of visuospatial attention to the visual cue, as the cue onset occurred long before (1 s) the AMI α analysis window and the visual cue was at a central fixation point, not co-localized with the target. Instead, as with absolute alpha power, alpha lateralization likely reflects inhibition of neural activity related to ignored stimuli, mediated by high alpha power in the hemisphere ipsilateral to the locus of attention (Jensen and Mazaheri, 2010, Wöstmann et al., 2016).

In the time window before a potential switch in talker, the alpha power was strongly lateralized in both continuous and switch trials (Fig. 5), reflecting suppression of the distraction CVs and selection of the auditory object in the attended direction (Kerlin et al., 2010). When the talker switched location in the second half of the trial, the hemispheric lateralization of alpha power was disrupted, but not when there was no switch. This may reflect spatial confusion: auditory selective attention may begin with allocating spatial attention and binding an auditory object to a location in space to assist in streaming (Kerlin et al., 2010). When a talker suddenly switches location, the system has to disassociate this auditory object with the location and associate the new talker with the target location. Our results thus appear to reflect the interactions between bottom-up discontinuity and top-down switching of attention (Desimone and Duncan, 1995). Future work should investigate this topographical pattern using imaging methods with higher spatial resolution (i.e., high-density EEG).

Task performance has been previously shown to relate to some variation of enhancement of N1 amplitudes (Choi et al., 2014) and change in alpha power during stimulus presenta-

tion (Kerlin et al., 2010, Wöstmann et al., 2015, 2016). However, we do not yet understand how the disruption of auditory attention is reflected in cortical responses, or how this relates to behavioral performance. Here, we find that the suppression of the N1 evoked by the CV following the switch in talker predicts the behavioral cost associated with the discontinuity (Fig. 6): a subject with a larger suppression of N1 shows a greater behavioral cost of the switch. We find a similar relationship with the ERD in the alpha band and behavioral performance: a larger desynchronization of alpha is associated with a larger decrease in behavioral performance. This pattern is inconsistent with previous work that shows that a larger ERD is associated with correct trials and better performance (Dimitrijevic et al., 2017). The changes in alpha power observed here presumably play a different role than in such previous tasks. Specifically, the ERD we report is induced involuntarily by talker discontinuity; it is not the result of a voluntary, top-down control of processing. Further investigation is needed to understand the generators and the many roles of alpha oscillations. It is also important to investigate whether similar effects (and of the same magnitude) are observed when the speaker switches to a new third speaker in the attended location rather than the two speakers flipping location, as was done in this study. It may be that the involuntary interruption of attention would be reduced. Regardless, we can conclude that the relative suppression of alpha and N1 caused by the perceptual discontinuity of the target talker limits one's ability to successfully attend to a sequence of syllables from a particular direction.

5. Conclusions

In summary, it is important not only to understand how cortical processing of attention enhances the sensory representation of sound mixtures, but also to understand the limitation of the system and when and how it fails. We show that perceptual discontinuities, which are common in acoustic settings, disrupt the neural mechanisms that facilitate sustained auditory spatial attention. The changes observed here demonstrate that talker continuity has an obligatory influence on selective auditory attention and affects listening in multi-source environments.

6. Acknowledgement

We would like to thank Jens Hjortkjær for his feedback on the manuscript. This work was supported by the H.C. Ørsted Foundation (Individual grant to: G.M.), the Oticon Centre of Excellence for Hearing and Speech Sciences (CHeSS), and NIH R01 DC013825 (to B.G.S.-C.).

7. References

- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J Mem. and Lang.*, 59(4):390–412.
- Best, V., Ozmeral, E., Kopco, N., and Shinn-Cunningham, B. (2008). Object continuity enhances selective auditory attention. *Proc Natl Acad Sci U S A*, 105(35):13174–13178.
- Best, V., Shinn-Cunningham, B., Ozmeral, E., and Kopco, N. (2010). Exploring the benefit of auditory spatial continuity. *J Acoust Soc Am*, 127(6):EL258–264.
- Bressler, S., Masud, S., Bharadwaj, H., and Shinn-Cunningham, B. (2014). Bottom-up influences of voice continuity in focusing selective auditory attention. *Psychological research*, 78(3):349–360.
- Chait, M., de Cheveigné, A., Poeppel, D., and Simon, J. Z. (2010). Neural dynamics of attending and ignoring in human auditory cortex. *Neuropsychologia*, 48(11):3262–3271.
- Choi, I., Rajaram, S., Varghese, L. A., and Shinn-Cunningham, B. G. (2013). Quantifying attentional modulation of auditory-evoked cortical responses from single-trial electroencephalography. *Frontiers in human neuroscience*, 7:115.
- Choi, I., Wang, L., Bharadwaj, H., and Shinn-Cunningham, B. (2014). Individual differences in attentional modulation of cortical responses correlate with selective attention performance. *Hearing research*, 314:10–19.
- Colby, C. L. and Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual review of neuroscience*, 22(1):319–349.
- Dai, L. and Shinn-Cunningham, B. G. (2016). Contributions of sensory coding and attentional control to individual differences in performance in spatial auditory selective attention tasks. *Frontiers in Human Neuroscience*, 10.
- Delorme, A. and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Meth*, 134(1):9–21.
- Desimone, R. and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu Rev Neurosci*, 18:193–222.
- Dimitrijevic, A., Smith, M. L., Kadis, D. S., and Moore, D. R. (2017). Cortical alpha oscillations predict speech intelligibility. *Frontiers in human neuroscience*, 11.

- 450 Ding, N. and Simon, J. Z. (2012). Emergence of neural encoding of auditory objects
451 while listening to competing speakers. *Proceedings of the National Academy of Sciences*,
452 109(29):11854–11859.
- 453 Dujardin, K., Derambure, P., Defebvre, L., Bourriez, J., Jacquesson, J., and Guieu, J. (1993).
454 Evaluation of event-related desynchronization (erd) during a recognition task: effect of
455 attention. *Electroencephalography and clinical neurophysiology*, 86(5):353–356.
- 456 Frey, J. N., Mainy, N., Lachaux, J.-P., Müller, N., Bertrand, O., and Weisz, N. (2014).
457 Selective modulation of auditory cortical alpha activity in an audiovisual spatial attention
458 task. *Journal of Neuroscience*, 34(19):6634–6639.
- 459 Hillyard, S. A., Hink, R. F., Schwent, V. L., and Picton, T. W. (1973). Electrical signs of
460 selective attention in the human brain. *Science*, 182(4108):177–180.
- 461 Huang, S., Chang, W.-T., Belliveau, J. W., Hämäläinen, M., and Ahveninen, J. (2014). Lat-
462 eralized parietotemporal oscillatory phase synchronization during auditory selective atten-
463 tion. *Neuroimage*, 86:461–469.
- 464 Jensen, O. and Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha
465 activity: gating by inhibition. *Frontiers in human neuroscience*, 4:186.
- 466 Kaiser, J., Lutzenberger, W., Preissl, H., Ackermann, H., and Birbaumer, N. (2000). Right-
467 hemisphere dominance for the processing of sound-source lateralization. *Journal of Neuro-*
468 *science*, 20(17):6631–6639.
- 469 Kenward, M. G. and Roger, J. H. (1997). Small sample inference for fixed effects from
470 restricted maximum likelihood. *Biometrics*, pages 983–997.
- 471 Kerlin, J. R., Shahin, A. J., and Miller, L. M. (2010). Attentional gain control of ongoing
472 cortical speech representations in a “cocktail party”. *Journal of Neuroscience*, 30(2):620–
473 628.
- 474 Klimesch, W. (1999). Eeg alpha and theta oscillations reflect cognitive and memory perfor-
475 mance: a review and analysis. *Brain research reviews*, 29(2):169–195.
- 476 Klimesch, W., Sauseng, P., and Hanslmayr, S. (2007). Eeg alpha oscillations: the inhibition-
477 timing hypothesis. *Brain research reviews*, 53(1):63–88.
- 478 Klimesch, W., Schack, B., Schabus, M., Doppelmayr, M., Gruber, W., and Sauseng, P. (2004).
479 Phase-locked alpha and theta oscillations generate the p1–n1 complex and are related to
480 memory performance. *Cognitive Brain Research*, 19(3):302–316.

- Lu, P. H., Lee, G. J., Raven, E. P., Tingus, K., Khoo, T., Thompson, P. M., and Bartzokis, G. (2011). Age-related slowing in cognitive processing speed is associated with myelin integrity in a very healthy elderly sample. *Journal of clinical and experimental neuropsychology*, 33(10):1059–1068.
- Maddox, R. K. and Shinn-Cunningham, B. G. (2012). Influence of task-relevant and task-irrelevant feature continuity on selective auditory attention. *Journal of the Association for Research in Otolaryngology*, 13(1):119–129.
- Michalka, S. W., Rosen, M. L., Kong, L., Shinn-Cunningham, B. G., and Somers, D. C. (2015). Auditory spatial coding flexibly recruits anterior, but not posterior, visuotopic parietal cortex. *Cerebral Cortex*, 26(3):1302–1308.
- Näätänen, R., Gaillard, A. W., and Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta psychologica*, 42(4):313–329.
- Nunez, P. L., Reid, L., and Bickford, R. G. (1978). The relationship of head size to alpha frequency with implications to a brain wave model. *Electroencephalography and clinical neurophysiology*, 44(3):344–352.
- Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). Fieldtrip: open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. *Computational intelligence and neuroscience*, 2011:1.
- Picton, T. and Hillyard, S. (1974). Human auditory evoked potentials. ii: Effects of attention. *Electroencephalography and clinical neurophysiology*, 36:191–200.
- Pinheiro, J. and Bates, D. (2000). *Mixed-effects models in S and S-PLUS*. Springer-Verlag, New York, NY.
- Polich, J. (1989). P300 from a passive auditory paradigm. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 74(4):312–320.
- Schaalje, B. G., McBride, J. B., and Fellingham, G. W. (2002). Adequacy of approximations to distributions of test statistics in complex mixed linear models. *J Agricult, Biol, Environ Stats.*, 7(4):512–524.
- Schneider, B. A. and Pichora-Fuller, M. K. (2001). Age-related changes in temporal processing: implications for speech perception. In *Seminars in hearing*, volume 22, pages 227–240. Copyright© 2001 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel.:+ 1 (212) 584-4662.
- Slepian, D. (1978). Prolate spheroidal wave functions, Fourier analysis, and uncertainty V: The discrete case. *Bell Syst Tech J*, 57(5):1371–1430.

- 514 Smith, D. V., Davis, B., Niu, K., Healy, E. W., Bonilha, L., Fridriksson, J., Morgan, P. S.,
515 and Rorden, C. (2010). Spatial attention evokes similar activation patterns for visual and
516 auditory stimuli. *Journal of cognitive neuroscience*, 22(2):347–361.
- 517 Strauß, A., Wöstmann, M., and Obleser, J. (2014). Cortical alpha oscillations as a tool for
518 auditory selective inhibition. *Front Hum Neurosci*, 8:350.
- 519 Thomson, D. (1982). Spectrum estimation and harmonic analysis. *Proc IEEE*, 70(9):1055–
520 1096.
- 521 Thut, G., Nietzel, A., Brandt, S. A., and Pascual-Leone, A. (2006). α -band electroencephalo-
522 graphic activity over occipital cortex indexes visuospatial attention bias and predicts visual
523 target detection. *Journal of Neuroscience*, 26(37):9494–9502.
- 524 Wöstmann, M., Herrmann, B., Maess, B., and Obleser, J. (2016). Spatiotemporal dynamics
525 of auditory attention synchronize with speech. *Proceedings of the National Academy of*
526 *Sciences*, 113(14):3873–3878.
- 527 Wöstmann, M., Herrmann, B., Wilsch, A., and Obleser, J. (2015). Neural alpha dynamics in
528 younger and older listeners reflect acoustic challenges and predictive benefits. *Journal of*
529 *Neuroscience*, 35(4):1458–1467.

530 Insert Supplementary Figure 1 here.

ACCEPTED MANUSCRIPT

Figure 1: (A) Trial design. Each trial started with a visual cue to indicate the side to be attended. The cue was followed by a fixation dot at the center of the screen, then the stimulus presentation. Following the stimulus, the response screen was shown, prompting the listener for a response. Feedback was provided on each trial. (B) Two streams of CV were presented on each trial, one spoken by a male and the other by a female speaker. The streams were separated using interaural time differences corresponding to approximately $\pm 30^\circ$. In the continuous trials, the talker at each location remained the same. In contrast, in the switch trials, the two talkers swapped locations in the third CV presentation. (C) The stimulus timing was designed to allow isolation of the ERPs for each CV. The trial began with a noise-burst, indicated in black, followed by the start of the leading/target stream. The lagging/masker stream began 0.18 s after the leading stream, creating an asynchrony in the CV onsets. The colored envelope superimposed on the plot represents the talker at that location. (D) Scalp topography of the N1 response to the first target CV. White circles indicate the electrodes used for ERP analysis.

Figure 2: (A) Behavioral performance for each condition. The black whisker plots show population results with horizontal lines indicating across-subject medians; error bars depict the maximum and minimum percent correct observed in each condition. Results for individual listeners are indicated by circles, with gray lines connecting results in the two conditions. *** $P < 0.001$. (B) Error rates as a function of target CV position in trials with only a single target.

Figure 3: (A) Grand average epoched EEG response for the active listening continuous (black) and switch (red) trials along with example topographies for each trial type. Vertical grey lines indicate the N1 of CVs in the leading/target stream, while the orange lines indicate the N1s of the CVs in the lagging/distractor stream. The yellow highlighted region indicates the time of the CVs following the switch in talkers, while the light blue highlighted region shows the time of the CVs after the switch. Topographies present the scalp distribution of N1 amplitude for the fourth CV in the leading stream in the to-be-attended continuous, and to-be-attended switch trials. (B) Grand average epoched EEG response for the passive continuous (dashed black) and switch (dashed red) trials. Topographies represent the scalp distribution of N1 amplitude for the third CV in the leading stream in the passive listening continuous and switch trials. (C) Average peak N1 amplitude across subjects for each CV in the target stream for the passive (open box) and active (filled box) conditions. A more negative value on the ordinate indicates a larger N1. Lines in each box plot indicate the median. Highlights correspond to the switch and post-switch CVs, as in panel A and B. * $P < 0.05$, ** $P < 0.01$.

Figure 4: Power in the alpha band, as a function of time, compared across conditions. The highlighted region in blue represents the time window in which the alpha power was significantly reduced in the switch trials relative to the continuous trials. $*P < 0.05$ after adjustment for multiple comparisons. Dashed lines indicate the onset of CVs in the target stream. The scalp topography of the average difference in alpha power between switch and continuous trials is shown on the right over the blue-highlighted time window where the difference reached statistical significance.

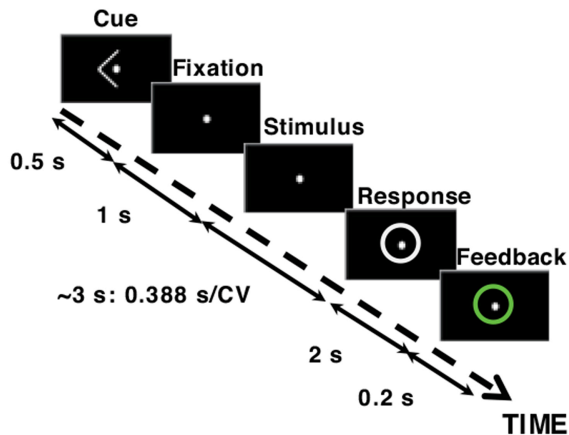
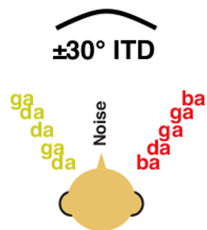
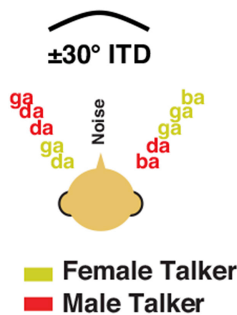
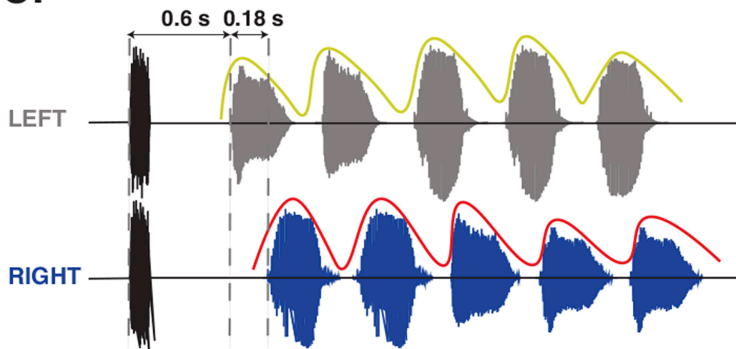
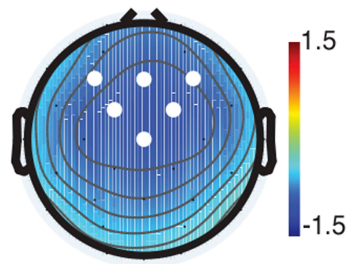
Figure 5: Topographic maps of the $AMI\alpha$ in two time periods (before and after a potential switch in talker) for continuous (A) and switch (B) trials. Bar graphs show mean across the posterior half of channels (excluding frontal channels) on the left hemisphere (LH) and right hemisphere (RH). Error bars indicate ± 1 SEM. $AMI\alpha$ showed a significant hemispheric lateralization (LH > RH) in both conditions before a potential switch. This lateralization remained significant in the second time window in the continuous trials where the talker remained in the same location (A: right panel). In contrast, when the talker switched location in the switch trials, the lateralization pattern was disrupted and was no longer significant. $*P < 0.05$; $**P < 0.01$; n.s., not significant.

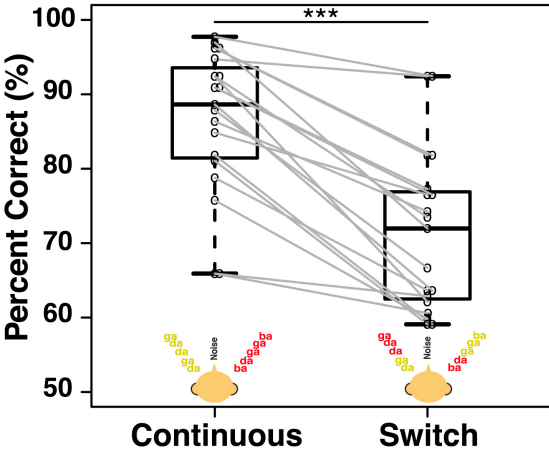
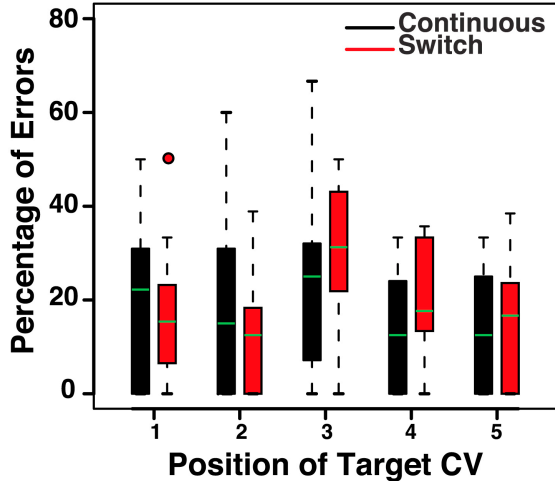
Figure 6: Relationship between the behavioral cost of talker discontinuity, defined as (% correct in Continuous - % correct in Switch trials), and (A) the difference in the N1 in continuous vs. switch (larger negative values indicate larger suppression of the N1 in the switch trials, corresponding to greater neural disruption of attention) and (B) the decrease in power in the alpha band, both calculated in a time window immediately following the switch in talker. Dashed lines represent 90% confidence intervals. $*P < 0.05$.

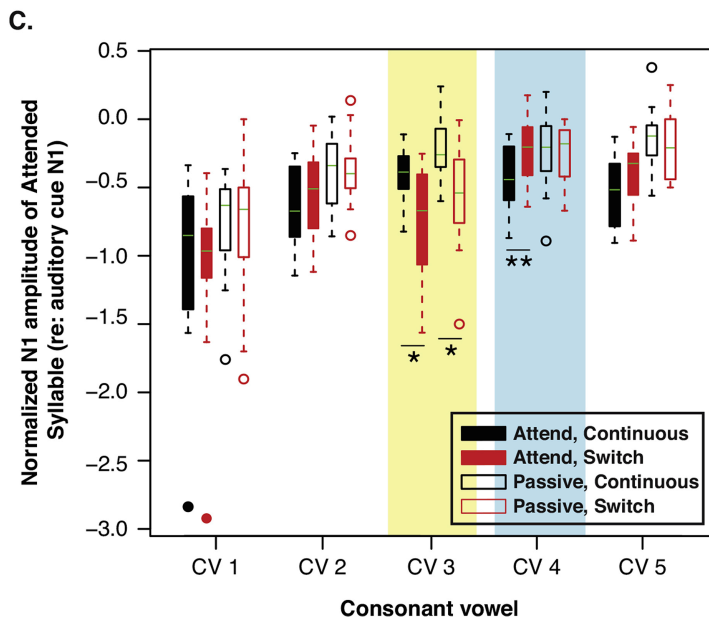
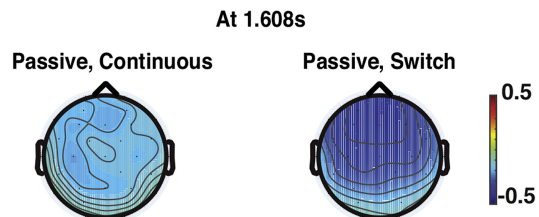
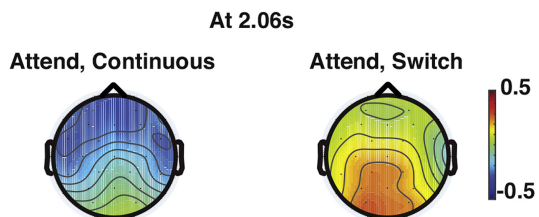
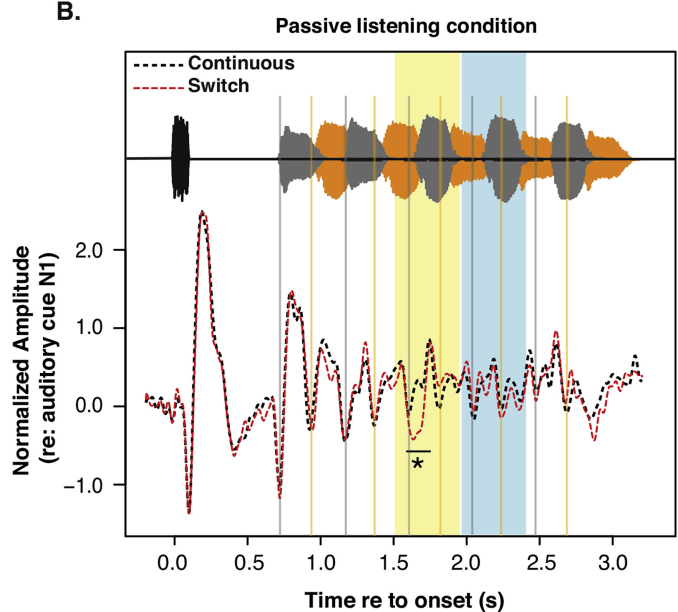
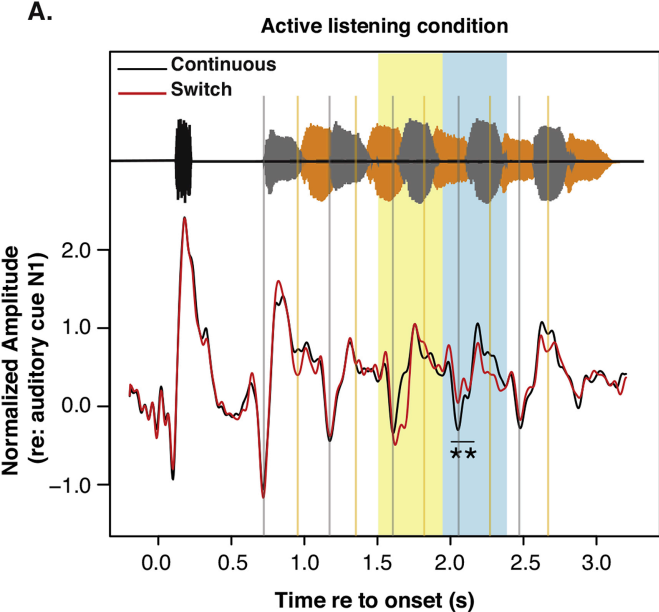
Figure S1: Power in the alpha band, as a function of time, when listeners are instructed to attend to a talker, regardless of location. The stimuli presented were the same as those in the data reported in the manuscript. In the switch trials, the talkers swapped location. The yellow highlighted region represents the time window in which the target and masker talker swap locations. Dashed lines indicate the onset of CVs in the target stream. We find no significant difference in alpha power between the continuous and switch trials, in contrast to when listeners are instructed to attend to a location (Fig. 4 in manuscript).

Table 1: Attentional modulation of N1 analysis, *p<0.05, **p<0.01

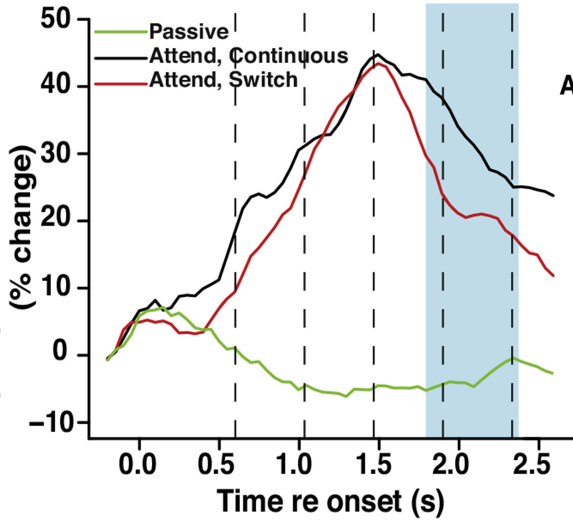
CV	Continuous trials, passive vs. active	Switch trials, passive vs. active
1	$z=-2.32^*$	$z=-1.7^*$
2	$z=-1.76^*$	$z=-1.4^*$
3	$z=-1.68^*$	$z=-1.03$
4	$z=-2.13^*$	$z=-0.23$
5	$z=-3.18^{**}$	$z=-1.75^*$

A.**B.****Continuous Trial****Switch Trial****C.****D.**

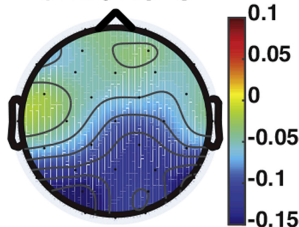
A.**B.**



Alpha power re to baseline

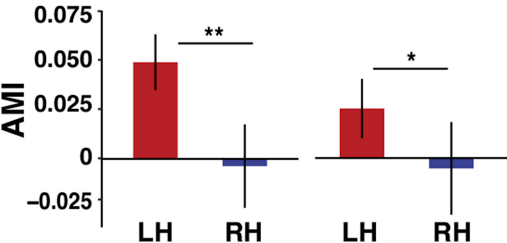
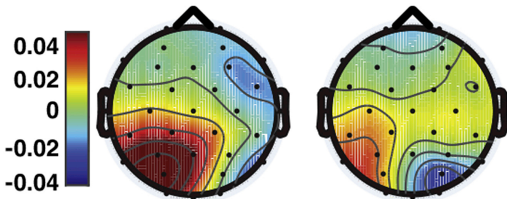


Alpha power difference
 $t=1.79-2.37$ s

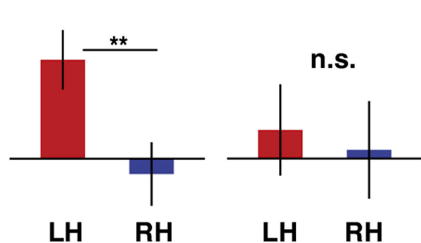
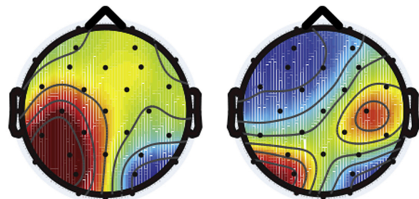


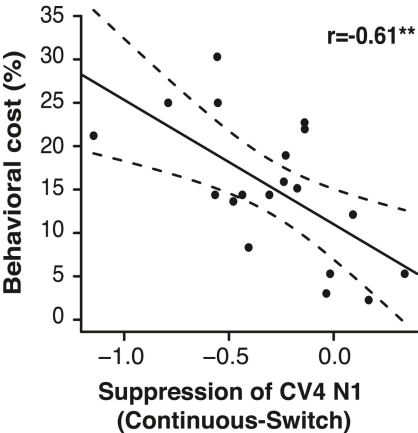
A.**Attend, Continuous**

Time: 0-1.466 s Time: 1.467-3.2 s

**B.****Attend, Switch**

Time: 0-1.466 s Time: 1.467-3.2 s



A.**B.**